ELSEVIER

Journal of Neuroscience Methods 69 (1996) 175–188

JOURNAL OF NEUROSCIENCE METHODS

Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability

Michale S. Fee, Partha P. Mitra, David Kleinfeld *

AT&T Bell Laboratories 600 Mountain Ave. Murray Hill, NJ 07974, USA

Received 5 October 1995; revised 11 March 1996; accepted 12 March 1996

Abstract

Neuronal noise sources and systematic variability in the shape of a spike limit the ability to sort multiple unit waveforms recorded from nervous tissue into their single neuron constituents. Here we present a procedure to efficiently sort spikes in the presence of noise that is anisotropic, i.e., dominated by particular frequencies, and whose amplitude distribution may be non-Gaussian, such as occurs when spike waveforms are a function of interspike interval. Our algorithm uses a hierarchical clustering scheme. First, multiple unit records are sorted into an overly large number of clusters by recursive bisection. Second, these clusters are progressively aggregated into a minimal set of putative single units based on both similarities of spike shape as well as the statistics of spike arrival times, such as imposed by the refractory period. We apply the algorithm to waveforms recorded with chronically implanted micro-wire stereotrodes from neocortex of behaving rat. Natural extensions of the algorithm may be used to cluster spike waveforms from records with many input channels, such as those obtained with tetrodes and multiple site optical techniques.

Keywords: Extracellular recording; Multi-unit spike train; Waveform variability; Stereotrode; Neocortex

1. Introduction

Studies of the response properties of neurons, as defined by the discharge of action potentials, have produced considerable advances in our understanding of how nervous systems code sensory and motor information. However, the mechanisms by which responses are formed and integrated within the brain remain to be elucidated. Essential to this pursuit is the ability to study, on a moment-to-moment basis, the responses and interactions of a large number of neurons as a means to characterize ensemble behavior and the detailed timing between spiking output of different neurons. Furthermore, economic constraints on the acquisition of neuronal data in behaving animals is such that one wishes to sample as many neurons as possible.

Extracellular recordings of brain activity often contain signals from more than one neuron. The decomposition of these multi-unit signals into the component action potentials is possible only if the shape of the waveforms from individual neurons differ at the location of the electrode. On the one hand, the shape of the extracellular signal generated by an action potential is fairly stereotypical in appearance throughout the mammalian brain, with a rapid positive going component that lasts a few hundred microseconds, followed by a slower negative going component that may last 2–3 times as long. On the other hand, extracellular action potential waveforms differ in detail as a consequence of the type and spatial distribution of currents in the cell and as a function of the position and geometry of the electrode. These differences provide a means to classify different waveforms as belonging to the same neuron. Extracellular sources of noise and intrinsic spike-to-spike variability obfuscate the classification process.

A number of algorithms have been developed to classify spike waveforms on the basis of differences in shape (for a review see Schmidt, 1984). These fall into two large categories: feature clustering and template matching. In feature clustering algorithms, a few properties of the waveform are measured, such as spike height, duration, and recovery rate. Many action potentials from the same neuron will tend to have similar properties, and will therefore form regions of high density, or clusters, in plots of the distribution of these properties. A human operator or a classification algorithm identifies and places boundaries

* Corresponding author. Department of Physics 0319, University of California, 9500 Gilman Drive, La Jolla, CA 92093, USA.

PII S0165-0270(96)00050-7

between different clusters of waveforms. The shortcoming of traditional feature clustering algorithms is that one does not necessarily know which features are optimal or how many independent features to incorporate. Thus, much information about the waveform is lost as waveforms from different neurons differ in subtle ways that may not be captured in the measured features.

Template matching algorithms classify on the basis of the overlap of the spike waveform with a set of previously determined template waveforms. They have the advantage that classification is based on the entire spike waveform. In the simplest implementations the template waveforms are found by visual inspection of many samples of the collected spikes (Bergman and DeLong, 1992), and thus suffer from biases and ill-defined selection criteria. The process of determining the templates is essentially that of modelling the observed set of spikes as a number of fixed average waveforms. In a recent, more sophisticated approach, the modeling of spike waveforms directly in the high-dimensional space of the sampled waveform has been automated (Lewicki, 1994). Single unit action potentials are modeled as the sum of an average waveform plus white noise with Gaussian distributed amplitudes. Using a soft clustering technique, a set of model waveforms is found that describes the set of waveforms observed on an electrode. Under the assumption of Gaussian noise, it is possible to define quantitatively the probability that a certain number of spike models describe the spikes in the observed signal. The final number of spike models may then be chosen as that with the highest probability, given the data. One serious drawback of this technique is that the waveform modeling, and thus the fidelity of the sorting, depends entirely on the restrictive assumptions that waveform noise is isotropic with Gaussian distributed amplitudes.

Here we present a new automatic procedure for spike classification in which no a priori assumptions are made about the distribution of spike waveforms within a single-unit cluster. The design of the algorithm was motivated by a recent analysis of spike waveforms recorded in rat neocortex (Mitra et al., 1995), which shows that waveform variability is highly anisotropic for all single-unit clusters observed. Furthermore, many waveforms have a large deterministic component of variability associated with the preceding interspike interval, resulting in temporal modes with non-Gaussian distributions. The present algorithm naturally encompasses such sources of variability, and thus may be particularly valuable for sorting waveforms from neurons that produce bursts of spikes.

2. Experimental methods

We record from layers II to VI of SI vibrissal cortex of Long-Evans rats. Four independently adjustable stereoelectrodes (McNaughton et al., 1983) are implanted through the intact dura mater into cortex. In brief, the electrodes are constructed from a twisted pair of 25 μm polyamide coated tungsten wires (California Fine Wire, Grover City, CA). The ends of the electrodes are cut with sharp scissors at a 45° angle and gold plated. Electrode impedances in physiological saline are typically 0.1 MΩ (1 kHz) for both reactive and resistive components, producing a Johnson noise of ~ 30 nV/$\sqrt{\text{Hz}}$ over the band of interest. Signals are amplified ($\times 10^4$), band-pass filtered between 300 Hz (5-pole Bessel high-pass filter) and 10 kHz (8-pole constant-phase low-pass filter), and digitized at 25 kHz using a spike acquisition program (Discovery; Datawave Technologies, Longmont, CO). The care and experimental manipulation of our animals are in strict accord with guidelines from the National Institutes of Health (1985) and have been reviewed and approved by our local institutional animal care and use committee.

Because the spatial separation of the two wires comprising the stereotrode is comparable to the size of neurons in the neocortex, one finds that the waveform from a single neuron is different on each wire. Thus, two neurons may have quite similar waveforms on one wire of the electrode, but quite different waveforms on the other wire, greatly improving the discriminability of action potentials. A threshold crossing of either signal of the stereotrode pair triggers the acquisition of the stereotrode spike waveforms on both wires; the threshold levels are independently adjustable at the time of data acquisition. Typically we set the thresholds so that some of the data collected is multi-unit activity that just exceeds the threshold.

When a threshold crossing is detected on either wire of a stereotrode pair, 32 samples of the waveform from each of the two wires of the electrode, which we denote $V_x(t)$ and $V_y(t)$, are saved. Both waveforms are centered with respect to the particular waveform that first crosses the threshold level. The voltage sample with the largest amplitude (waveform peak) is set as the fifth sample in the stored waveform; the companion waveform is shifted in register. A time stamp saved with each waveform indicates the time of the waveform peak with a 100 μs resolution. A scatter plot of the peak values of $V_x(t)$ and $V_y(t)$ is shown in the inset of Fig. 4a; the exclusion produced by the threshold settings results in a lack of spikes in the lower-left corner of the distribution. The 32 waveform samples from each wire of the stereotrode define a 64-dimensional vector, denoted $\vec{V}$. We refer to this vector as the spike waveform-pair. A number of such waveform-pairs define a set of points in a 64-dimensional space.

3. Hierarchical clustering

The need for a new approach to spike classification is made apparent by two aspects of the spike waveform variability observed in rat neocortical neurons (Mitra et al., 1995). First, we have found that spikes from individual neurons form highly anisotropic clusters in the space of waveforms. The largest directions of variability within a

cluster are not generally parallel to the line that connects the means of two clusters. This poses a priori difficulties for either hard or soft clustering schemes, which use single means (and possibly a variance) to describe individual clusters. The difficulty is illustrated by the two clusters in Fig. 1a, which shows the result of representing each cluster by a single mean and using hard clustering. The hyperplane normal to the line joining the means is the usual classification boundary and evidently results in misclassification of spikes. The problem is that the clusters have more structure than can be captured in a single mean waveform and a single variance. The internal structure of a cluster may be parameterized in one of several ways. We choose to do so by describing each cluster by several means. As illustrated in Fig. 1b, if we cluster the data set into four means, the boundary between the two single-unit clusters is well described by the boundaries between regions 1 and 3, and regions 2 and 4.

A second aspect of spike waveform variability is that some major components of variability are not well described by a Gaussian distribution, and in general, do not appear to have a simple analytical form. In particular, we observe systematic changes in spike shape that are determined by the length of the last interspike interval. Further, waveform changes due to slow electrode drift can result in a non-Gaussian distributions of waveforms around the mean of a single-unit cluster.

Given the anisotropic and non-Gaussian nature of single-unit clusters, modeling spike waveforms under the assumption of Gaussian mixtures is likely to give inaccurate results, particularly for the purpose of estimating the number of underlying waveforms in a data set. Such an approach is likely to assign two or more distinct model

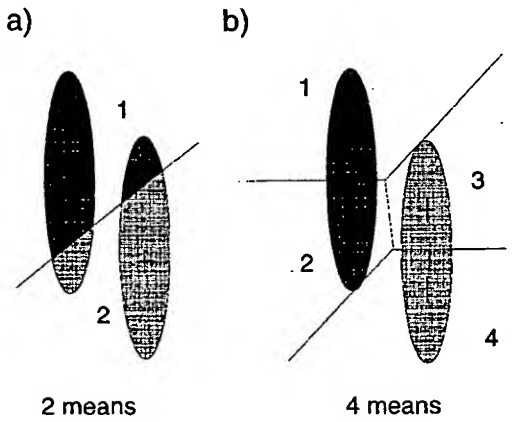

Fig. 1. The geometrical effect of waveform variability on clustering of spike waveforms. (a) The two shaded areas represent the anisotropic distribution of two different spike waveforms in some planar projection. If each of the anisotropic clusters is represented by a single mean, the naive choice of cluster boundary produces unsatisfactory separation of the clusters. (b) If four means are used to represent the same two clusters, then each of the four regions produced contains spikes from only one cluster. The regions must then be combined appropriately, i.e., regions 1 and 2 define a single cluster.

waveforms to a single anisotropic cluster. We conclude, therefore, that the modeling of spike waveforms should be independent of detailed a priori assumptions about waveform distributions.

Spikes can be classified based on waveform similarity, however, this is not always adequate given the observed large variation of spike waveform shape within a cluster. Additional information for classifying spikes is present in the spike arrival times, as the biophysical properties of neurons lead to stereotypical properties of the interspike interval (ISI) distribution. The most characteristic of these is a refractory period that prevents a single neuron from spiking twice within a period of 1–2 ms (for neocortical cells see McCormick et al., 1985). Intrinsic neuronal dynamics can also produce a characteristic burst of action potentials (Connors and Gutnick, 1990) with a corresponding peak in the ISI distribution. Finally, significant correlations exist between the spike trains of different neurons (Fetz et al., 1991), although the magnitude of the cross-correlation coefficient is typically small. We have incorporated in our algorithm information from spike timing statistics to identify clusters of waveforms likely belonging to different neurons.

The algorithm is a three-step procedure. In the first step, all of the spike waveform pairs are centered in time to minimize jitter in the position of the peak as an additional source of waveform variability. In the second step, the centered waveforms are sorted into a large number of clusters, typically 10-times the expected number of single-unit clusters, by recursive bisection of an initial sampling of waveforms. All spike waveforms are then classified into these initial clusters. Clusters with few elements are discarded. At this point, the problem of clustering the entire data set, typically 100 000 spikes in our examples, is reduced to one of aggregating an initial set of clusters, typically 40 in our examples, into a final set of clusters, each of which contains spike waveforms from the same putative single-unit. We automate the aggregation by the use of: (i) a measure of the similarities of waveform shape between clusters; and (ii) a measure of the interspike intervals between the arrival times of spikes in different clusters.

We give a detailed description of each step in the algorithm. A flow chart summarizing these steps is shown in Fig. 2.

3.1. Centering waveforms

The discrete sampling of the spike waveforms leads to a relative timing jitter of one sample period in the onset time of spikes in a cluster. Since the rise time of spike waveforms is comparable to the sampling period, the timing jitter contributes significantly to the waveform variability. Each waveform is therefore resampled and centered with a resolution much greater than the rise time. We use a composite measure of spike time found from the center of

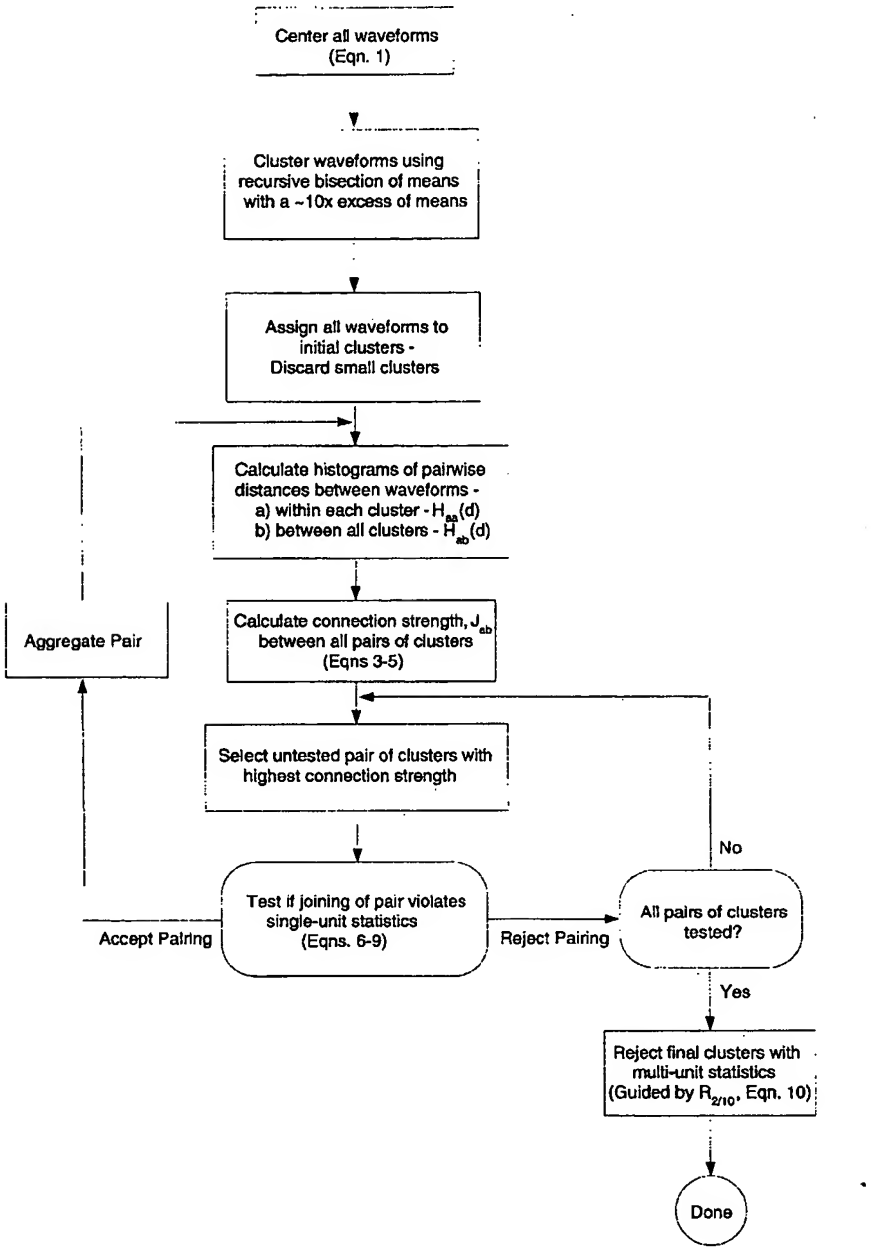

Fig. 2. Flow chart of automatic spike classification algorithm.

mass of the sum of the waveforms on the stereotrode pair. The center of mass defines a peak time, T_p, and is computed only over the samples that lie within the peak, i.e.,

$$T_p = \frac{\sum_{i=2}^{6} (i-4)\big(V_x(i) + V_y(i)\big)}{\sum_{i=2}^{6} \big(V_x(i) + V_y(i)\big)} \quad (1)$$

where $V_x(i)$ and $V_y(i)$ are the i-th sample of the x and y channels of the stereotrode. The primary advantage of this formulation is that if a particular cluster has a large waveform on one wire and a small waveform on the other, the large waveform contributes most to the center of mass. The waveform is resampled by cubic spline interpolation (Press et al., 1990) to place the peak time, T_p, at some fixed sample number in the resampled waveform [1].

[1] We have compared the center-of-mass procedure for centering waveforms to that using an explicit model for the entire waveform (Lewicki, 1994) and found the spike times calculated by the two procedures to have a high degree of correlation ($\rho = 0.95$) for a typical waveform cluster. The difference between these two measures of spike time is distributed with a half-width of less than 0.05 sample period, indicating the utility of our center-of-mass procedure for model-independent waveform centering.

3.2. Initial clustering

A subset of spikes from the data set [2] are selected for the initial clustering. The spikes are clustered into many means, typically 10-times the expected number of single-unit clusters in the data set, by recursive bisection of the data. The first bisection is performed as follows: (i) The mean of the data subset is found. (ii) A second mean is constructed by duplicating the first one and adding a component of random noise whose magnitude is small (~ 0.001) compared with the size of the clusters. (iii) Each vector in the subset is classified sequentially as belonging to the closest mean. Furthermore, the two means are continuously updated by the vectors assigned to them with a weight that decreases inversely with the number of vectors assigned. To reduce fluctuations during the initial few assignments, the means are not updated until a sufficient number of assignments, typically 5, have been made. (iv) The assignment procedure and the updating of means, steps (iii) and (iv), are repeated a few times (~ 3) to ensure that the means have settled to their equilibrium values. Subsequent bisections proceed as follows: (1) Every mean is duplicated, in analogy with step (ii) above, to form a new set of means with twice the number elements. (2) The assignment of all vectors to the new set of means and the update of the means proceeds as in steps (iii) and (iv) above. The bisection procedure stops after a predetermined number of bisections; we typically use 5 or 6 bisections.

Once the means have been found from the sample of spikes, all of the waveforms in the data set are classified with these means. During this classification procedure, the means are continuously updated by each assigned waveform vector [3]. We generally find a wide distribution of the number of spikes assigned to the initial clusters. Those with few waveforms are difficult to classify. Prior to aggregating the clusters, we discard all clusters whose size falls below a set threshold; a satisfactory choice of threshold is 0.01-times the data set size.

The outcome of the initial clustering procedure is shown in Fig. 4a. A sample of 2000 spike waveforms collected from a stereotrode were bisected into 64 clusters. The assignment into these clusters was carried out on the first 5000 waveforms of the same data set. The 5000 spike waveforms are shown projected into a plane selected to highlight the separation of the single-unit clusters of spikes. The means of the 38 largest initial clusters are indicated by the superimposed pluses.

3.3. Aggregation procedure

At this point the original waveform-pairs have been divided into a number of clusters. The remainder of the algorithm considers the aggregation of these clusters into the more global structure of single-unit clusters. The aggregation is carried out using, first, a measure of connectivity between pairs of clusters that is based on the distribution of pairwise distances between the waveforms that comprise each cluster. A large value of this measure indicates that two clusters contain spikes that are similar in shape and thus should be combined. A second input to the aggregation is provided by the distribution of interspike intervals. Our working hypothesis is that an appropriate refractory period characterizes spikes arising from a single unit. Thus, pairs of clusters should not be combined if that would significantly degrade any refractory period present in either of the original clusters.

The procedure by which these criteria are used to aggregate clusters is as follows (see Fig. 2): (i) Evaluate the matrix of connection strengths, $J_{ab} = J_{ba}$, between pairs of clusters a and b. (ii) Rank order the connection strengths and examine the most strongly connected pair of clusters. (iii) Test if the combination of these two clusters significantly degrades the refractory period that is present in the original clusters. (iv_a) If the test is not significant then combine the clusters and return to step one to a new matrix of connection strengths J_{ab}; note that each time a pair of clusters are aggregated, the rank of this matrix is decreased by one. (iv_b) If the test is significant, do not combine the clusters. Rather, examine the next most strongly connected pair of clusters, determined by J_{ab}, and repeat step three. (v) When all possible pairs of clusters have been rejected, the algorithm is terminated.

3.4. Connection strength between clusters

Two clusters may be considered strongly connected if they have, by some measure, a large number of points near the boundary between them. We quantify the connectivity by analogy with the energy at the interface of two groups of neighboring physical particles that interact via a short-range potential, U. The underlying idea is that two clusters that originate from the same single-unit will have a large value for the interface energy, whereas well separated clusters have few waveforms at their interface and will have negligible interface energy. This approach is superior to the use of the Euclidean distance between clusters means as measure of the connectivity, e.g., means 2 and 3

[2] The size of the subset of spikes chosen is a compromise between computational efficiency and the ability to sample the structure of the data. We have found that 2000 waveform-pairs are adequate for the recursive bisection described here.

[3] The continuous update of the means during waveform assignments allows for potential slow drifts in the spike waveform due to electrode movement. A small weighting on the order of 0.005 is sufficient for our data sets; waveform drift is only occasionally seen over the course of our 30 min data sets, and then only for the largest (and therefore usually the best isolated) waveforms, for which the consequences of electrode movement are the least severe.

of Fig. 1 are closest in the Euclidean sense yet belong to separate single-unit clusters.

We define the interface energy between clusters a and b as

$$E_{ab} = \sum_{j=1}^{N_a} \sum_{k=1}^{N_b} U\left(|\vec{V}_j^a - \vec{V}_k^b|\right) \quad (2)$$

where $\vec{V}_j^a$ is the j-th vector in cluster a, $\vec{V}_k^b$ is the k-th vector in cluster b; and $|\vec{V}_j^a - \vec{V}_k^b| = \sqrt{\sum_{i=1}^{64} \left[V_j^a(i) - V_k^b(i)\right]^2}$ is the distance between the vectors. The potential U is chosen, ad hoc, to be

$$U(d) = \exp(-d/d_0) \quad (3)$$

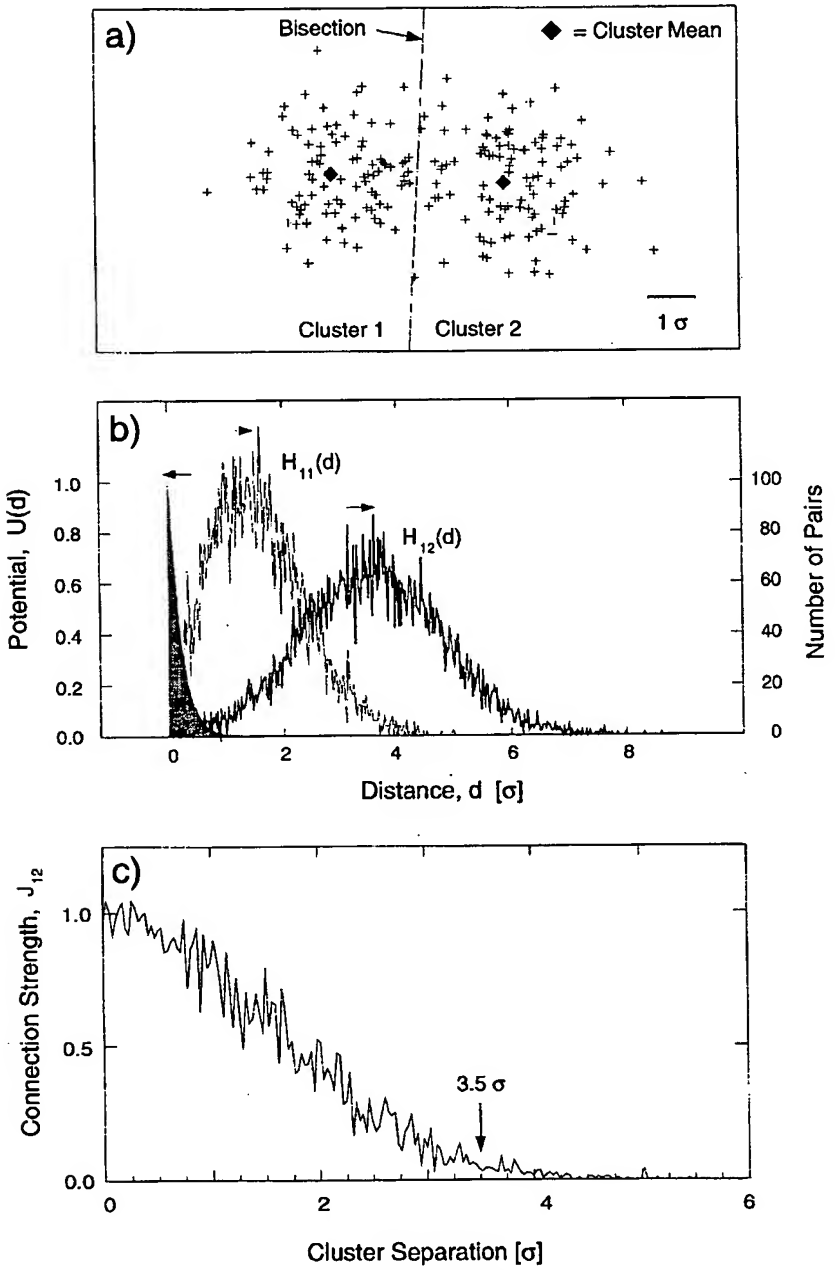

Fig. 3. Illustration of the calculation of connection strength used in the spike sorting algorithm (Eqs. (3)–(5)). (a) Random vectors (+), representing spike waveforms, are generated from two Gaussian distributions with means spaced 3.5σ apart. The clustering algorithm described in the text (Section 3.2) is used to partition the total distribution into two clusters whose means are shown (diamonds) near the middle of the distributions on either side of the boundary. (b) Histograms of the pairwise distances between all pairs of vectors within cluster 1 ($H_{11}(d_i)$) and between all pairs of vectors between the two clusters ($H_{12}(d_i)$). Also shown (shaded curve) is a plot of the potential, $U(d)$ (Eq. (3)), with $d_0 = 0.1\sigma$. (c) Plot of the connection strength as a function of the separation of the means of the two Gaussian distributions. The arrow indicates the value of J_{12} for the separation of the means used in (a): $J_{12} \approx 0.05$.

where d_0 is a scale parameter; it is approx. 0.1-times the average distance between waveform-pairs within a typical cluster.

The energy E_{ab} depends only on the distribution of the set of pairwise distances $\{|\vec{V}_j^a - \vec{V}_k^b| j = 1 \ldots N_a, k = 1 \ldots N_b\}$. We approximate this distribution as a coarse-grained histogram, $H_{ab}(d_i)$, where d_i is an index spanning the full range of distances. The total energy is now approximated by

$$E_{ab} \cong \sum_{d_i=0}^{d_{max}} U(d_i) H_{ab}(d_i) \quad (4)$$

Similar expressions define E_{aa} and E_{bb}. The energy E_{ab} is

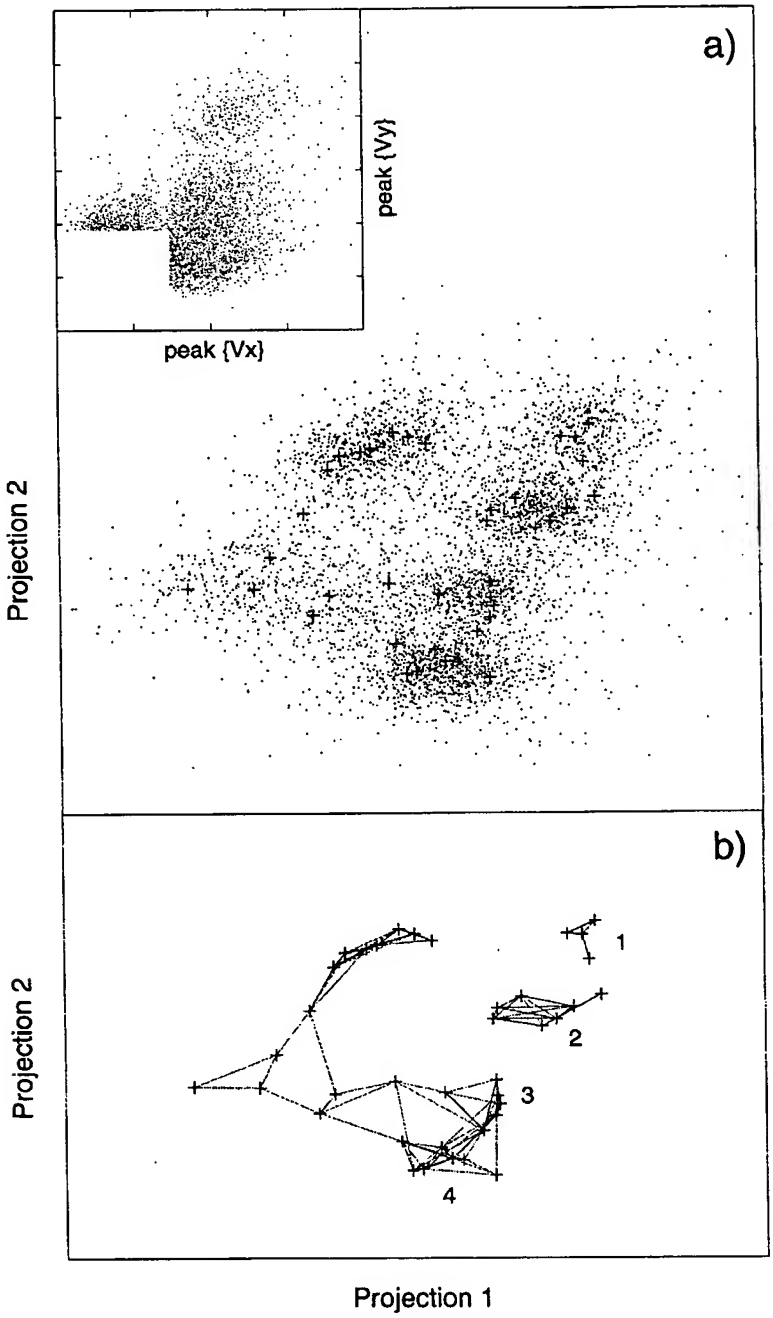

Fig. 4. (a) Scatter plot of a set of 5000 acquired stereotrode waveforms. The projection used was chosen to highlight the separation of most of the clusters. The spikes were clustered into 64 means using the recursive bisection algorithm (Section 3.2). The positions of the 38 largest means are shown as superimposed pluses. Both axes are drawn to the same scale. The inset is a scatter plot of the same data projected into the coordinates defined by the peak values of the two stereotrode waveforms, V_x and V_y. (b) The 38 means from (a) are reproduced with the lines between them representing the connectivity between the sub-clusters. Darker lines represent a higher connection strength.

scaled with respect to the intra-cluster interaction energies E_{aa} and E_{bb} to define a connectivity between clusters a and b, denoted J_{ab}, i.e.

$$J_{ab} = 2\frac{E_{ab}}{E_{aa} + E_{bb}} \tag{5}$$

This is the measure of connectivity we use in the aggregation process.

The pairwise aggregation of clusters requires that a new set of connection strengths must be considered at every iteration of the algorithm. The pairwise distance histograms $H_{ab}(d_i)$ for an arbitrary combination of clusters can be expressed in terms of the histograms for the original set of clusters. For example, if clusters a and b are combined to form a new cluster c, the intra-cluster distance histogram is given by $H_{cc}(d) = H_{aa}(d) + H_{bb}(d) + H_{ab}(d)$. The inter-cluster distance histogram to a third cluster d is given by $H_{cd}(d) = H_{ad}(d) + H_{bd}(d)$. In our implementation, the pairwise distances are placed into 500 bins spanning the range of distances seen in our data sets; from zero to roughly 3% of the maximum distance possible with 64 samples of 12-bit integers. An additional step taken to reduce computation time is that pairwise distances are computed only for a fixed number of spikes in each cluster, regardless of the number of spikes assigned. We find that 100 spikes adequately represents the distribution and geometry of a cluster, consistent with the low dimension of the variability within a cluster (Fig. 8 and Section 5).

To examine the performance of the above measure of connectivity (Eqs. (3)–(5)) between clusters of known distributions, we consider the simple case of a mixture of two identical Gaussian distributions in two dimensions that are offset along some axis. 100 vectors were sampled from a mixture of two distributions (in two dimensions) separated by 3.5σ along the x-axis, as shown in Fig. 3a. The vectors are clustered with two means as described (Section 3.2). The resulting means are shown as diamonds at roughly the center of each underlying distribution and the vertical line passing through the center is the bounding line between the two resulting clusters (Fig. 3a).

The distribution of pairwise distances between the vectors assigned to cluster 1, $H_{11}(d_i)$, and the distribution of distances between all pairs of vectors containing one vector from each cluster $H_{12}(d_i)$, are shown in Fig. 3b. One can see that the distribution $H_{12}(d_i)$ is strongly skewed toward longer distances, as expected. From these distributions, we calculate the interface energy E_{12}, the intra-cluster energy E_{11} and E_{22} (Eqs. (3)–(5)) and finally, the connection strength J_{12} (Eq. (5)). For the cluster separation shown in the example, we find that J_{12} is small, ~ 0.05, consistent with the clusters being fairly well separated. In general, J_{12} is a strong function of the separation distance through the short range potential (Eq. (3)), as shown in Fig. 3c. Although the exact shape of $H_{12}(d_i)$ will depend on the distribution of vectors, taken as Gaussian in this example, the form of $H_{12}(d_i)$ at small distances depends only on the density of vectors at the interface between the two clusters. Thus our measure of the connection strength is fairly independent of the shape of the clusters.

3.5. Inclusion of ISI statistics

We now consider the distribution of time intervals that separate spikes in one cluster, a, from an immediately preceding spike in another cluster, b. If the two clusters represent the same unit, then the distribution of the inter-cluster spike intervals (ICSIs) should have as good a refractory period as the ISI distributions of the parent clusters. We statistically test if the ICSI distributions exceed either of the ISI distributions within either of the clusters during the refractory period. If there is such an excess, then we prevent the clusters from combining. The details of the test used are given below.

For purposes of the statistical test, we use the conditional distribution of the ICSIs and the ISIs where the intervals are those smaller than some maximum time. Since we are interested in the short time behavior of the distributions, we take the maximum time to be 10 ms. The task is to test whether the ICSI distribution exceeds either of the ISI distributions in the 0–2 ms time interval. A crude way of achieving this goal would be to use a straightforward Chi-square test. Since data is relatively sparse in this small interval, we use a more powerful test involving the empirical distribution functions of the spike intervals. The empirical distribution function $F_{ab}(\tau)$ of the ICSIs between clusters a and b is defined as

$$F_{ab}(\tau) = m_{ab}(\tau)/M_{ab} \tag{6}$$

where $m_{ab}(\tau)$ is the number of ICSIs between cluster a and b whose absolute value is less than τ, and M_{ab} is the total number of ICSIs being considered. Since we are studying the conditional distribution of ISIs and ICSIs shorter than 10 ms, M_{ab} is the number of ICSIs between clusters a and b less than 10 ms long. The corresponding quantities $F_a(\tau)$ and $F_b(\tau)$ are defined analogously on the set of ISIs. Consider the case where we test if $F_{ab}(\tau)$ significantly exceeds $F_a(\tau)$ in the interval $\tau < \tau_1 = 2$ ms. This can be achieved by means of the scaled difference function

$$S_a = (\tau) = \sqrt{\frac{(M_{ab}M_a)}{(M_{ab} + M_a)}}\,[F_{ab}(\tau) - F_a(\tau)] \tag{7}$$

and the associated statistic

$$D_a(\tau) = \max_{\tau < \tau_1}[S_a(\tau)] \tag{8}$$

.The quantity $D_b(\tau)$ is defined analogously.

The above is a slight generalization of the usual one-

sided Kolmogorov-Smirnoff statistic used in non-parametric comparisons of empirical distribution functions (Stephens, 1986). The maximum is taken over a sub-interval of the range over which the distribution is calculated. Let τ be distributed over the interval (τ_{min},τ_{max}), and $D_b(\tau)$ defined above is calculated for $\tau_1 = \tau_{min} + f \cdot (\tau_{max} - \tau_{min})$. It can be shown (Mitra, unpublished), under the null hypothesis, that the underlying distributions are identical on the sub-interval (τ_{min},τ_1), that the probability distribution function for $D_a(\tau_1)$ is

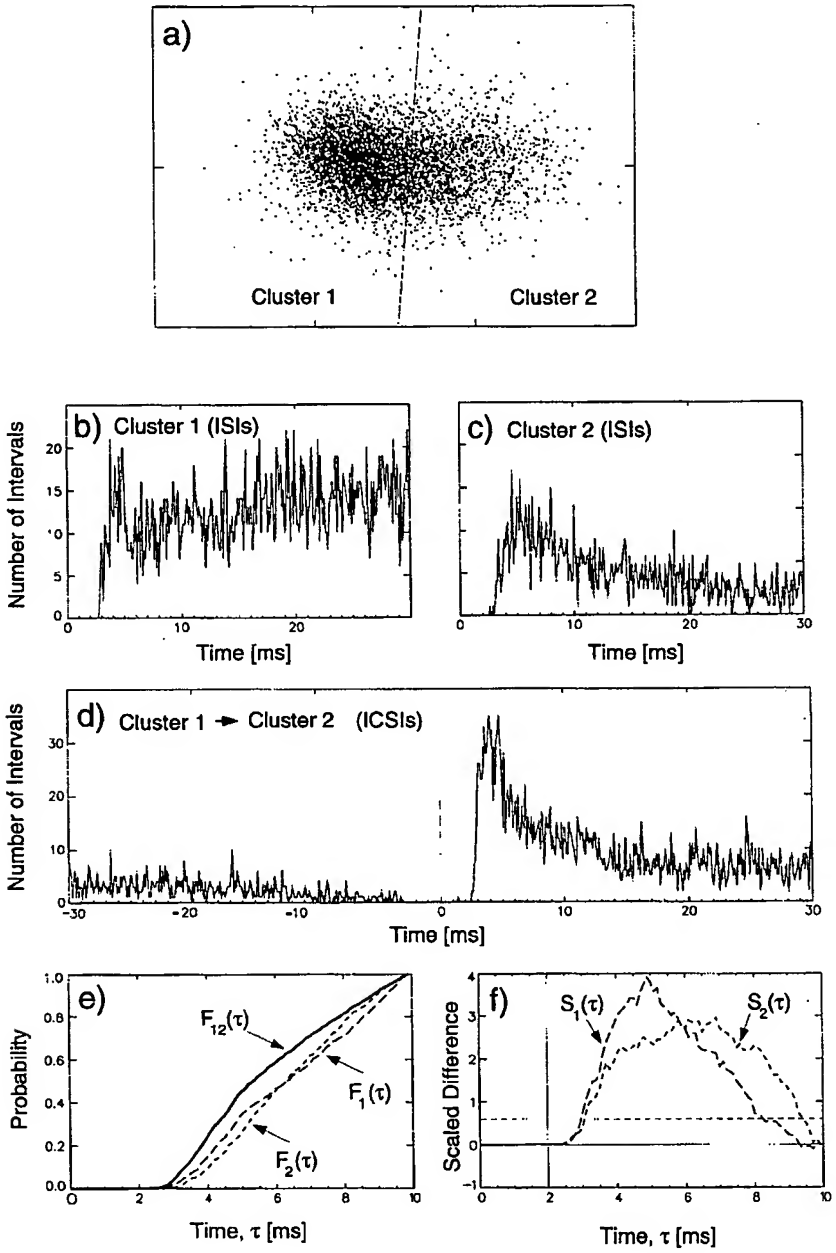

Fig. 5. Illustration of the use of spike statistics for the aggregation of two clusters. (a) Planar projection of two clusters obtained near end of the aggregation process; the plane passes through the mean of each cluster. The connection strength between the clusters is $J_{12} \approx 0.14$ (Eqs. (3)–(5)). (b,c) Histograms of inter-spike intervals (ISIs) of clusters 1 and 2, respectively. Note the complete suppression of intervals less than 2 ms, indicating single-unit spike trains. (d) Histogram of the intra-cluster spike intervals (ICSIs). Negative intervals begin with a spike from cluster 2 and end with a spike from cluster 1, with no intervening spikes. Note the complete suppression of ICSIs less than 2 ms, indicating that these two clusters are probably samples from the same single-unit cluster. Also note the strong peak between 3 and 10 ms, indicating that spikes in cluster 2 tend to closely follow spikes in cluster 1. (e) Probability distributions of the absolute values of the ICSIs ($F_{12}(\tau)$) and the distributions of the ISIs of cluster 1 ($F_1(\tau)$) and cluster 2 ($F_2(\tau)$) (Eq. (6)). (f) Scaled difference functions, $S_1(\tau)$ and $S_2(\tau)$ (Eq. (7)), indicate no significant deviations of the ICSI distribution from the ISI distribution of each sub-cluster within the 0–2 ms window. The 95% confidence interval for significant deviations within the 0–2 ms interval is shown (horizontal dashed line).

$$P(D_a < \lambda) = \frac{1}{2}\left[1 + \mathrm{Erf}\left(\frac{\lambda}{\sqrt{2f(1-f)}}\right)\right] - \frac{1}{2}\exp(-2\lambda^2)\left[1 - \mathrm{Erf}\left(\frac{2(1-2f)\lambda}{\sqrt{2f(1-f)}}\right)\right] \quad (9)$$

where $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\mathrm{d}t$. In the present instance $\tau_{min} = 1.2$ ms; the shortest interval between spikes allowed by the acquisition software, $\tau_{max} = 10$ ms, $\tau_1 = 2$ ms, and $f = 0.091$. The 95% confidence interval for the above distribution is given by $D_a = 0.63$, i.e. $P(D_a < 0.63) = 0.95$. We reject the null hypothesis, that is the ICSI distribution function is equal to the ISI distribution function of either cluster for periods less than the chosen refractory period, if the statistic computed above exceeds 0.63. In this case, the clusters are not allowed to combine.

To illustrate the above procedure for incorporating spike statistics into the cluster aggregation process, two clusters that originate from a single neuron exhibiting a large variation in spike waveform associated with short interspike intervals (Fee et al., 1995) are shown in Fig. 5a. The plane of projection passes through the means of each cluster. Although the two clusters do not appear to be well separated, the connection strength between them is only $J_{ab} = 0.14$ (Eqs. (3)–(5)). As seen in Fig. 5b and c, the ISI histograms of the individual clusters exhibit a complete suppression of intervals less than 2 ms, indicating that they each represent a single-unit spike train. The histogram of ICSIs is shown in Fig. 5d, where the intervals are negative if the interval begins with a spike from cluster 2. There are two relevant features. First, there is a complete suppression of ICSIs less than 2 ms, indicating that these two clusters are probably samples of the same single-unit spike train. Secondly, there is a strong peak between 3 and 10 ms, indicating that spikes in cluster 2 tend to closely follow spikes in cluster 1.

The distribution functions of the ICSIs, $F_{ab}(\tau)$, and the ISIs, $F_a(\tau)$ and $F_b(\tau)$, over the interval 0–10 ms, are shown in Fig. 5e. The scaled difference functions, $S_a(\tau)$ and $S_b(\tau)$ (Eq. (7)), are shown in Fig. 5f. The maxima of the difference functions in the interval 0–2 ms are both much smaller than the 95% confidence interval, shown as a horizontal dashed line. Although it is not clear whether the two clusters should be assigned to a single cluster on the basis of waveform information alone, the statistical test of the refractory period clearly indicates that they should be combined to form a single-unit cluster.

3.6. Determining single-unit clusters

The pairwise aggregation (Fig. 2) proceeds until all of the original clusters are subsumed into a final, relatively small set of clusters, as illustrated by the dendrogram in Fig. 6. At each step in the procedure the pair of clusters with the largest connection strength (Eq. (5)) is combined, so long as the fidelity of the interspike interval statistics of the aggregate is not degraded relative to that of the original clusters (Eq. (8)). The aggregation stops when the statistical test precludes additional combinations. The final set of clusters represent both putative single units and multiple unit activity. The categorization of a cluster as a putative single unit is based on the fidelity of the refractory period for the unit. We compare the number of interspike intervals that lie between 0 and 2.0 ms, which encompasses the refractory period, with those that lie within a relatively long interval. The precise length of the long interval, here as in our statistical criteria for vetoing the aggregation of clusters, depends on the neuronal dynamics in so far as some neurons exhibit a tendency to spike in bursts or, conversely, exhibit suppressed firing at modest intervals.

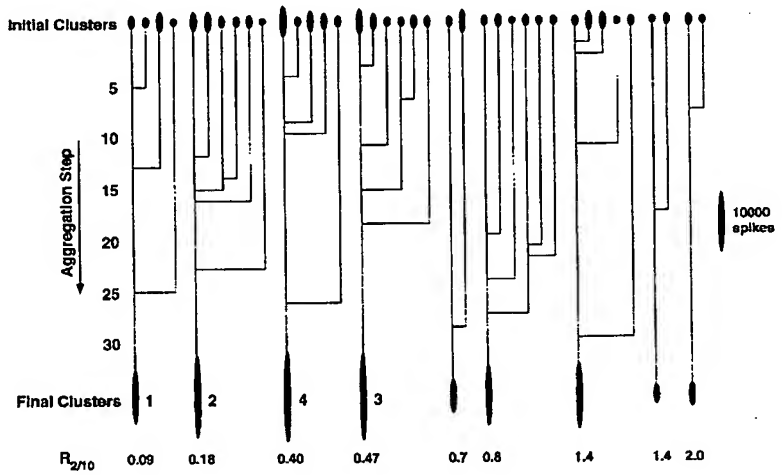

Fig. 6. Dendrogram of the pairwise aggregation process for the data set of Fig. 4. The initial clusters are represented at the top; the height of the ellipse represents the number of spikes in the cluster. Recombinations are sequential in the downward direction in the figure. The first four final clusters are those labelled in Fig. 4. The ratio $R_{2/10}$ represents the depth of the refractory period, and therefore the single unit quality of the cluster (Eq. (10)).

As a simple measure of the fidelity of the refractory period, we compared the number of ISIs in the 1.2–2.0 ms interval [4] with those in the 1.2–10.0 ms interval in terms of the normalized ratio:

$$R_{2/10} = \frac{8.8}{0.8} \times \frac{F(2)}{F(10)} \qquad (10)$$

where $F(\tau)$ is the distribution function of ISIs, defined by Eq. (6). Qualitatively, a value of $R_{2/10}$ near zero suggests that the unit has a well defined refractory period, while a value $R_{2/10} \sim 1.0$ clearly implies that the waveforms represent multiple units.

4. Results

We focus on the data set shown in Fig. 4. In 30 min, roughly 10^5 spike waveforms were collected from a single stereotrode (Section 2). The clustering algorithm was run on an Intel 80486-DX2/66 based laboratory computer running IDL scientific programming language (Research Systems Inc., Boulder CO). The time to classify a set of spikes into single-units breaks down as follows. The time required to center the waveforms and perform cluster assignments is linear in the size of the data set, and ran at 250 and 500 spikes per s, respectively. The times for the initial clustering and for the pairwise distance calculations are non-linear in the number of spikes, and, as a practical matter, are carried out on a fixed sized sample of waveforms; the initial clustering carried out on 2000 spike waveforms took 5 min and pairwise distance calculations on roughly 3800 spikes took 40 min. The aggregation of the 38 clusters to produce the 9 final clusters took approx. 5 min. The results of the aggregation process are summarized in Fig. 6.

The waveforms and autocorrelation histograms for 4 of the 9 final clusters for this data set are shown to be single units. in Fig. 7. We identify clusters 1 and 2, with the values $R_{2/10} = 0.09$ and $R_{2/10} = 0.18$, as single units. The other two clusters, with $R_{2/10} = 0.40$ and $R_{2/10} = 0.47$, show varying degrees of multi-unit contamination. The remainder of the final clusters depicted in Fig. 6, but omitted from Fig. 7, are clearly multi-unit in nature. Cross-correlation analysis of these clusters (not shown) indicate that they are not components of a single bursting neuron.

The algorithm presented here has been used to classify neuronal waveforms from roughly 100 sets of stereotrode data. Typically, 2–3 single-unit clusters, as identified by auto-correlation analysis, are isolated on each stereotrode implanted into deep layers of neocortex. Often several

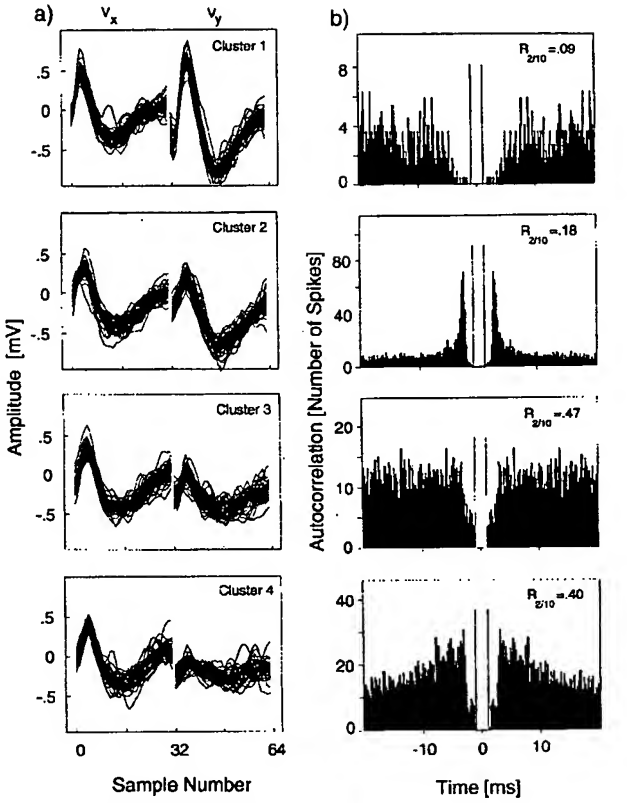

Fig. 7. (a) Representative spike waveforms for 4 of the 9 final clusters identified in the data set of Fig. 6. The waveform recorded on each wire of the stereotrode pair is shown. (b) Corresponding autocorrelation functions. The bin size is 0.1 ms.

more multi-unit clusters may also be classified. From a sample of 220 final clusters from 44 sets of data, clusters with $R_{2/10} < 0.20$ constituted about 25% of the sample [5]. These always had autocorrelation functions that smoothly approach zero at small times, consistent with their classification as single-units. A small fraction of clusters, < 10%, had intermediate values of $R_{2/10}$, i.e. $0.2 < R_{2/10} < 0.40$. For these clusters it is necessary to inspect the autocorrelation function in order to establish the fidelity of the refractory period. Lastly, the value of $R_{2/10}$ for the major fraction of clusters exceeded 0.4; for these clusters the autocorrelation functions were always indicative of multiple units.

The anisotropy of the single-unit clusters may be demonstrated by considering the spectrum of variability associated with the principal components of the waveform-pairs comprising the clusters. We consider one

[4] Again, the low end of this interval is determined by the spike acquisition system, which will not accept spikes following an interval less than 1.2 ms at the given digitization rate.

[5] The number of final clusters and the distribution of $R_{2/10}$ will depend to a large extent on the settings of the spike threshold during spike acquisition. The more sensitive the thresholds are set, the larger the fraction of collected spikes associated with multi-unit clusters will be. Over a wide range of threshold settings, the single-unit clusters and the associated values of $R_{2/10}$ are insensitive to the threshold.

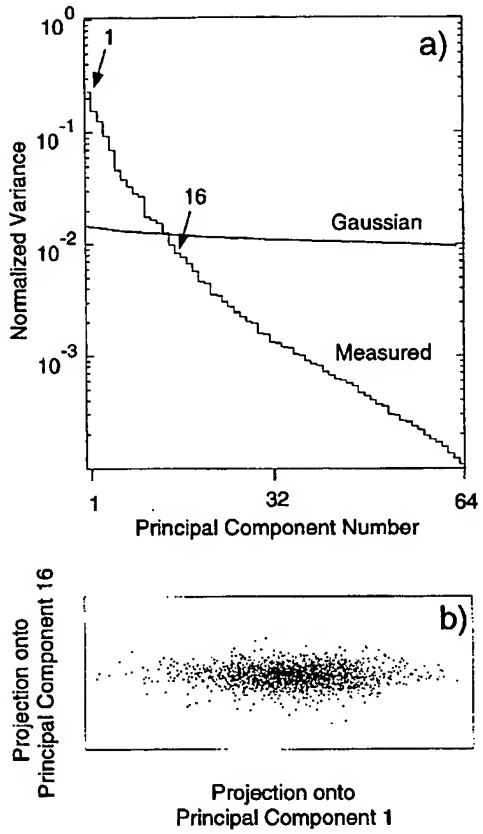

Fig. 8. Results of principal components analysis of the spike waveform residuals. (a) Eigenvalue spectrum of the covariance matrix found by principal component analysis of ~10000 spike waveforms. The eigenvalues, which indicate the contribution of variance from each component, are sorted and normalized so the sum over all elements is 1.0. Also plotted is the sorted eigenvalue spectrum of residuals of Gaussian distributed white noise with the same number of samples. (b) Projection of the spike waveforms into the plane given by the first and fifteenth principal components. The scale along each axis is the same.

cluster (cluster 1 of Fig. 7) and calculate the eigenvectors and eigenvalues of the covariance matrix of the residuals of the spike waveforms; the eigenvectors form the principal components and the eigenvalues denote the variance associated with each component (Golub and Kahan, 1965; Golomb et al., 1993). The spectrum of sorted eigenvalues decreases rapidly as a function of the principal component number, as shown in Fig. 8a. This indicates that the distribution is highly anisotropic, with 5 principal components accounting for half of the total variance. Another measure of anisotropy is that the variance along the largest principal components is more than three orders of magnitude larger than the variance along the smallest principal component. For isotropic Gaussian noise, all eigenvalues are statistically identical; the sorted eigenvalue spectrum for a random cluster with the same number of spikes is also shown in Fig. 8a. The anisotropy is clear in the projection of spike waveforms on a plane defined by the first and sixteenth principal components (Fig. 8b).

5. Discussion

We have described a general procedure to sort voltage waveforms into clusters of putative single units. There are three distinguishing features of our procedure and its implementation. First, we make no a priori assumption about the form of the variability in waveforms, allowing us to effectively sort spike waveforms into putative single-unit clusters in the presence of realistic distributions of neuronal noise. Second, we incorporate physiological restrictions on the statistics of spike arrival times directly into the algorithm. This is particularly valuable for the identification of waveforms from a putative single unit that produces bursts of spikes that may differ significantly in shape. Finally, each final cluster is described by a number of mean waveforms, rather than one, to account for intracluster structure.

The algorithm contains a number of free parameters. Some of these parameters control the size of sub-samples of the data, and are chosen to be large enough to appropriately sample the underlying structure in the data, but small enough to be computationally manageable. Some of the parameters could be defined as dimensionless quantities to ensure a degree of invariance over different data sets. For example, the scale of the potential $U(d)$ (Eq. (3)), could be determined for each data set based on the noise level, although we have not found this to be necessary. In our implementation of the algorithm, the values of all parameters were chosen on the basis of a careful examination of the performance of the algorithm on ~10 data sets. The performance of the algorithm appears insensitive to small changes in the values of the parameters.

In the present work we use an algorithm for the initial clustering of the waveforms that requires that we set, by hand, the initial number of clusters. More sophisticated algorithms may use a cost function to set this number. In the particular case of continuous waveforms, as opposed to the segmented waveforms used in this work, the algorithm described by Lewicki (1994) may be suitable for the initial clustering. For this purpose, Lewicki's algorithm has the desired property that it will produce an over abundance of clusters for anisotropic distributions of waveforms.

We do not address at present the decomposition of voltage waveforms that result from the overlap of neuronal waveforms. This omission is justified by the small fraction of waveforms in our data that are overlapped within the 1.2 ms interval over which the spike waveforms are sampled. On the other hand, some waveforms have a significant amplitude beyond this interval. The overlap of subsequent waveforms with this tail region leads to an asymmetric suppression of the ICSIs between two clusters for about 5% of our single units; this suppression is confined to the interval between +1.2 and +1.6 ms. The tests for a refractory period (Eqs. (6)–(9)) are relatively insensitive to this suppression. In other experimental situations, where the overlap of spike waveforms with the tail region of

previous spikes is problematic, there are ways to resolve this problem. First, with continuously acquired spike data, spike models may be subtracted from the continuous record to remove the contribution of the waveform tail (Lewicki, 1994). Second, if only segments of data containing spikes are acquired, the effects of such overlaps may be reduced by additional high-pass filtering of the spike waveforms.

With continuous records, it will be possible to implement a more sophisticated overlap decomposition algorithm such as that described by Lewicki (1994). Given the relatively small fraction of overlapped spike waveforms in most electrophysiology recordings, it does not appear to be necessary, or even desirable, to solve the classification problem and the overlap problem simultaneously. Rather, it should be satisfactory to implement our classification algorithm as a first step to find the sub-clusters associated with each single-unit cluster. A representation of spike overlap models may then be constructed using the means of those sub-clusters. In a second pass, the spike data would be tested against the overlap models. Given that the clusters cannot be assumed to be isotropic and Gaussian distributed, the Bayesian approach is likely to be unreliable in estimating the number of spikes contributing to the overlap. On the other hand, it seems that limiting spike overlap models to two waveforms should reduce the loss of spike events to a negligibly small level, except under conditions of extreme neuronal synchronization.

In the absence of a priori knowledge of spike waveforms, the entire clustering procedure is carried out in the original high dimensional space of the sampled time points (in our case $d = 64$). In general, however, the average spike waveforms of different neurons may be well represented in a lower dimensional space. To determine a set of vectors that span this space, we applied principal components analysis to a sample of 22 single-unit clusters. These were obtained from recordings throughout vibrissal cortex of rat (see Section 2) and included both regular- and fast-spiking waveforms (Simons, 1978). The first six principal components account for 99% of the power in the sample, as shown in Fig. 9a. Fig. 9b shows the time course of the first six modes. This result shows that spike waveforms of different neurons can be represented in a relatively low-dimensional space, as first shown by Abeles and colleagues (Abeles and Goldstein, 1977; Gerstein et al., 1983). Thus, clustering may be carried out using only the first 5–10 projections. The components of spike waveform variability not contributing to the discrimination between clusters are eliminated, resulting in a less noisy estimation of connection strengths. The dominant components correspond to the distinguishing 'features' among the waveforms. Lastly, we note that the large degree of correlation among the time-points in the components (Fig. 9a) indicates significant information will be discarded by spike classification methods based on measures of spike waveform derived from only a few time-points.

Several advantages may be had by projecting the spike

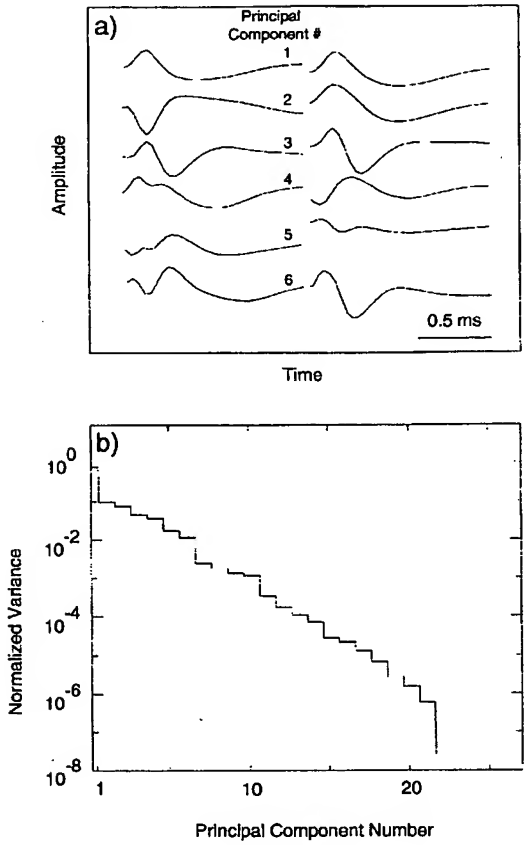

Fig. 9. Variability of average stereotrode waveforms from a sample of 22 isolated single units recorded throughout primary somatosensory cortex. (a) The first six principal components of the distribution of mean waveforms. (b) The eigenvalue spectrum of the covariance matrix shows the relative contribution of the principal components. The spectrum falls rapidly with mode number; the first six components account for roughly 99% of the power in the sample of average spike waveforms.

waveforms into the set of principal components described above, which we denote the spike basis. Foremost, the projections onto the spike basis set may serve as a filter for acceptable spike waveforms. The ratio of power in the first 20 modes to the power in the higher modes serves as a figure of merit for the quality of the spike waveform. We find that a ratio threshold of 20 eliminates the roughly 20% of the data set that appears to consist of multi-unit signals close to threshold, as well as the small percentage consisting of overlapping spike waveforms.

A final issue is the generality of our spike sorting procedure for the classification of extracellular data not acquired with stereotrodes. In our algorithm the separate waveforms of the stereotrode pair are concatenated in a single vector and, from a formal point of view, serve only to increase the dimensionality of the space of waveforms. As such, the algorithm can be trivially extended to encompass different degrees of freedom, e.g., single electrodes and tetrodes (Wilson and McNaughton, 1994). The extension to the analysis of records with large numbers of simultaneously acquired channels, such as those acquired

with arrays of surface electrodes (Pine, 1980; Meister et al., 1994) or with multiple site optical techniques (London et al., 1987; Parsons et al., 1991) is fundamentally straightforward but may present computational challenges in light of the large size of the vectors involved. In cases where a given neuron only contributes to the output of a small fraction of the total number of sensors, it will be useful to first extract the relevant spatial (or probe number) and temporal modes of the data by singular value decomposition, thus reducing the size of the data set without loss of information.

Acknowledgements

We thank R.A. Stepnoski for assistance with the acquisition software, and M. Lewicki and H.S. Seung for useful discussions.

References

Abeles, M. and Goldstein, H.M. (1977) Multispike train analysis, Proc. IEEE, 65: 762–772.

Bergman, H. and DeLong, M.R. (1992) A personal computer based spike detector and sorter: implementation and evaluation, J. Neurosci Methods, 41: 187–97.

Connors, B.W. and Gutnick, M.J. 1990. Intrinsic firing patterns of diverse neocortical neurons, Trends Neurosci., 13: 99–104.

Fee, M.S. et al. (1995) PLEASE PROVIDE FULL DETAILS.

Fetz, E., Toyama, K. and Smith, W. (1991) Synaptic interactions between cortical neurons. In A. Peters and E.G. Jones (Eds.), Cerebral Cortex, Vol. 9, Normal and Altered States of Function, Plenum Press, New York.

Gerstein, G.L., Bloom, M., Espinosa, L., Evanczuk, X. and Turner, M. (1983) Design of a laboratory for multi-neuron studies, IEEE Trans. Sys. Man. Cybern., 13: 668–676.

Golomb, D., Kleinfeld, D., Reid, R.C., Shapley, R.M. and Shraiman, B. (1993) On temporal codes and the spatiotemporal response of neurons in the lateral geniculate nucleus, J. Neurophysiol., 72: 2990–3003..

Golub, G.H. and Kahan, W. (1965) Calculating the singular values and pseudo-inverse of a matrix, SIAM Num. Anal., 2: 202–224.

Lewicki, M.S. (1994) Bayesian modeling and classification of neural signals, Neural Comput., 6: 1005–1030.

London, J.A., Zecevic, D. and Cohen, L.B. (1987) Simultaneous optical recording of activity from many neurons during feeding in Navanax, J. Neurosci., 7: 649–661.

McCormick, D.A., Connors, B.W., Lighthall, J.W. and Prince, D.A. (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex, J. Neurophysiol., 54: 782–806.

McNaughton, B.L., O'Keefe, J. and Barnes, C.A. (1983) The stereotrode: A new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records, J. Neurosci. Methods, 8: 391–397.

Meister, M., Pine, J. and Baylor, D.A. (1994) Multi-neuronal signals from the retina: acquisition and analysis, J. Neurosci. Methods, 51: 95–106.

Mitra, P.P., Fee, M.S. and Kleinfeld, D. (1995) The variability of extracellular spike waveforms is not random: A new algorithm for spike sorting, Soc. Neurosci. Abstr., 21: 115.

National Institutes of Health (1985) Guide for the Care and Use of Laboratory Animals, NIH Publication 85–23, Bethesda, MD.

Parsons, T.D., Salzberg, B.M., Obaid, A.L. Raccuia-Behling, F. and Kleinfeld, D. (1991) Long-term optical recording of patterns of electrical activity in ensembles of cultured Aplysia neurons, J. Neurophys., 66: 316–334.

Pine, J. (1980) Recording action potentials from cultured neurons with extracellular microcircuit electrodes, J. Neurosci. Methods, 2: 19–21.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., (1990) Numerical Recipes, Cambridge University Press, Cambridge.

Schmidt, E.M. (1984) Computer separation of multi-unit neuroelectric data: A review, J. Neurosci. Methods, 12: 95–111.

Simons, D.J. (1978) Response properties of vibrissal units in rat SI somatosensory neocortex, J. Neurophysiol., 41: 798–820.

Stephens, M. (1986) In R.B. D'Agostino and M. Stephens (Eds.), Goodness-of-Fit Techniques, Dekker, New York.

Wilson, M. and McNaughton, B.L. (1994) Dynamics of the hippocampal ensemble code for space, Science, 261: 1055–1058.